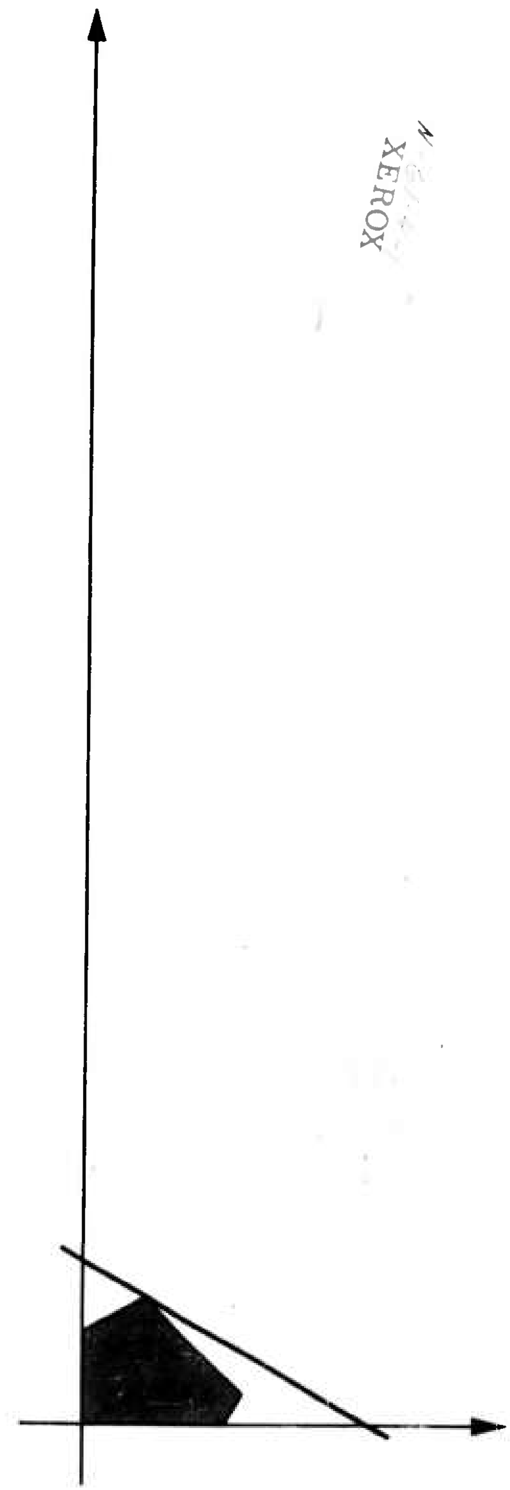
RESEARCH REPORT 2
14 APRIL 1961
I.E.R. 172–3

QUADRATIC PROGRAMMING

A VARIANT OF THE WOLFE-MARKOWITZ ALGORITHMS

by

George B. Dantzig



OPERATIONS RESEARCH CENTER

INSTITUTE OF ENGINEERING RESEARCH

UNIVERSITY OF CALIFORNIA - BERKELEY

QUADRATIC PROGRAMMING
A VARIANT OF THE WOLFE-MARKOWITZ ALGORITHMS

by

George B. Dantzig
Operations Research Center
University of California, Berkeley

April 14, 1961 Research Report 2

This research was supported in part by the Office of Naval Research under contract Nonr-222(83) with the University of California. Reproduction in whole or in part, is permitted for any purpose of the United States Government.

Quadratic Programming

A Variant of the Wolfe-Markowitz Algorithms

Although a convex quadratic objective can be treated by general convex programming, and also can be reduced to the convex separable case by a change of variables, the linear nature of its partial derivatives has given rise to an elegant theory important in its own light. It is doubtful at this writing that full potentiality of this theory has been realized.

Barankin and Dorfman (1958)* first pointed out that if the linear Lagrangian conditions of optimality were combined with those of the original system, the optimum solution was a basic solution in the enlarged system with the property that only one of certain pairs of variables were in the basic set. Markowitz (1956), on the other hand, showed that it was possible to modify the enlarged system and then parametrically generate a class of basic solutions with the above special property which converges to the optimum in a finite number of iterations. Finally, Wolfe (1959) proved, in an elegant way, that an easy way to do this is to modify the simplex algorithm so as not to allow a variable to enter the basic set if its "complementary" variable is already in the basic set. Thus by modifying a few instructions in a simplex code for linear programs it was possible to solve a convex quadratic program! We shall present here a variant of Wolfe's procedure. The chief difference is a tighter selection rule that results in a monotonically decreasing objective instead of a decreasing measure of "dual" infeasibility. It is believed to be computationally more efficient because there can be a greater decrease in the value of the quadratic function in each iteration.

---

* The name and bracketed date refers to references at the end of the report. Other references on quadratic programs are listed there also.

Quadratic programs can arise in several ways. Wolfe lists four in his paper as follows:

Regression: To find the best least-square fit to given data, where certain parameters are known a priori to satisfy inequality constraints.

Efficient Production: Maximization of profit, assuming linear production functions and linearly varying marginal costs; see Dorfman (1951).

Minimum Variance: To find the solution of a linear program with variable cost coefficients which will have given expected costs and minimum variance; see Markowitz (1959).

Convex Programming: To find the minimum of a general convex function under linear constraints and quadratic approximation; see White, Johnson and Dantzig (1958).

PRELIMINARIES:

Before stating the problem, let us note that every quadratic form can be conveniently expressed in terms of a symmetric matrix associated with its coefficients. For example, for n = 3 variables,

$$(1)\quad Q(x) = c_{11}x_1^2 + c_{22}x_2^2 + c_{33}x_3^2 + 2c_{12}x_1x_2 + 2c_{23}x_2x_3 + 2c_{13}x_1x_3$$

$$= [x_1,\ x_2,\ x_3] \begin{bmatrix} c_{11} & c_{12} & c_{13} \\ c_{12} & c_{22} & c_{23} \\ c_{13} & c_{23} & c_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = x^T C x$$

where T stands for transpose.

Definition: A quadratic form is called positive definite if $x^TCx > 0$ for all $x \neq 0$; it is called positive semi-definite if $x^TCx \geqslant 0$ for all x.

PROBLEM: <u>Find</u> $x = (x_1, x_2, \ldots, x_n) \geqslant 0$ <u>and</u> Min $Q(x)$ <u>satisfying</u>

$$(2) \qquad Ax = b \;, \qquad A = [a_{ij}] \;, \qquad i = 1, 2, \ldots, m$$

$$x^T C x = Q(x) \;, \qquad C = [c_{kj}] \;, \qquad k, j = 1, 2, \ldots, n$$

<u>where</u> $Q(x)$ <u>is positive semi-definite</u>.*

<u>Kuhn-Tucker Optimality Conditions</u>:** Let A_j, C_j denote the j^{th} column of A and C and let

$$(3) \qquad y_j = C_j^T x - \pi A_j \;, \qquad (\pi = \pi_1, \pi_2, \ldots, \pi_m) \;.$$

THEOREM 1: <u>A solution</u> $x = x^o$ <u>is minimal if there exists a</u> $\pi = \pi^o$, $y = y^o$ <u>such that, for</u> $j = 1, 2, \ldots, n$,

$$(4) \qquad Ax^o = b, \; x^o \geqslant 0 \;, \qquad \text{(Primal feasibility)},$$

$$(5) \qquad y_j^o = C_j^T x^o - \pi^o A_j \geqslant 0 \;, \qquad \text{("Dual" feasibility)},$$

$$(6) \qquad y_j^o \cdot x_j^o = 0 \qquad \text{(Complementarity)} \;.$$

PROOF: Rewrite $Q(x)$ in the form

$$(7) \qquad Q(x) - Q(x^o) = 2 \sum_{j=1}^{n} (C_j^T x^o)(x_j - x_j^o) + (x - x^o)^T C (x - x^o) \;.$$

In general, let x and x^o be any solutions satisfying $Ax = b$, then

$$(8) \qquad A(x - x^o) = \sum_{j=1}^{n} A_j (x_j - x_j^o) = 0 \;.$$

---

* If desired the theory is easily extended to include the addition of linear terms to $Q(x)$.

** Theorem 1 is, as indicated earlier, well known; we reprove it because it sets the stage for the development that follows.

Multiplying on the left by $2\pi^o$ and subtracting from (7) yields, for any $Ax = b$, $Ax^o = b$,

$$(9) \qquad Q(x) - Q(x^o) = 2 \sum_{j=1}^{n} (C_j^T x^o - \pi^o A_j)(x_j - x_j^o) + (x - x^o)^T C(x - x^o)$$

$$= 2 \sum_{j=1}^{n} y_j^o (x_j - x_j^o) + (x - x^o)^T C(x - x^o) \ ,$$

where y_j^o is defined by (3) for $x = x^o$. If in addition complementarity holds, $x_j^o \cdot y_j^o = 0$, then (9) simplifies to

$$(10) \qquad Q(x) - Q(x^o) = 2 \sum_{y_j^o \neq 0} y_j^o x_j + (x - x^o)^T C(x - x^o) \quad .$$

Finally, if primal and dual feasibility holds so that $x_j^o \geqslant 0$, $x_j \geqslant 0$, $y_j^o \geqslant 0$, then all terms in (10) are non-negative, therefore $Q(x) \geqslant Q(x^o)$.

Improving a Non-Optimal Solution. Consider the system

$$(11) \qquad \begin{aligned} &Ax &= b \ , \qquad x \geqslant 0 \ , \\ &Cx - A^T \pi^T - I_n y = 0 &, \qquad (I_n\text{: Identity Matrix}), \end{aligned}$$

where $x^T Cx$ is assumed to be positive semi-definite. Let x^o, π^o, y^o be a basic feasible solution associated with a basic set with the complementarity property; namely, for each j either x_j or y_j, but not both, are in the basic set. We shall assume further that the right hand side has been perturbed to insure that all basic solutions are nondegenerate. Note that neither π nor y are sign restricted; only $x \geqslant 0$ is required for a "feasible" solution to (11); an optimal solution will have been obtained if $y_j \geqslant 0$ and $x_j \cdot y_j = 0$ holds for all j.

THEOREM 2: If a basis is complementary and $y_s^o < 0$, then any increase of the nonbasic variable x_s, with adjustment of the basic variables, generates a class of solutions x', π', y', such that x^TCx decreases as long as $y_s' < 0$.

PROOF: Let x be any solution in the class generated by x_s and let x' be generated by $x_s = x_s'$. From (9), $Q(x) - Q(x') = 2y_s'(x_s - x_s') + (x - x')^T C(x - x')$ since for all $j \neq s$ either x_j or $y_j = 0$. The adjusted values of the basic variables are linear functions of x_s, hence it follows that $x - x' = (x_s - x_s')v$ where v is a constant vector. Hence, $Q(x) - Q(x') = (x_s - x_s')\left[2y_s' + (x_s - x_s')(v^TCv)\right]$ and it is clear that if $y_s' < 0$, the right hand side is negative for sufficiently small $(x_s - x_s') > 0$. Moreover for $Q(x)$ to decrease with an increase of $x_s \geqslant 0$ from x_s' to x_s'', it must be accompanied by $y_s' < y_s''$ because $Q(x'') - Q(x') = 2(x_s'' - x_s')y_s' + (x_s'' - x_s')^2 v^TCv =$
$= 2(x_s'' - x_s')y_s'' - (x_x'' - x_s')^2 v^TCv$ whence $2(y_s'' - y_s') = (x_s'' - x_s')v^TCv \geqslant 0$. But $v^TCv \neq 0$ because $v^TCv = 0$ implies for positive semi-definite forms $Cv = 0$ and $Q(x'') - Q(x') =$
$= 2(x_s'' - x_s')x'Cv + (x_s'' - x_s')^2 vCv = 0$ whereas $Q(x'') - Q(x') < 0$; hence $y_s'' > y_s'$.

THEOREM 3: If x_r drops as basic variable, introduction of y_r either causes x^TCx to decrease (and x_{r_1} or y_s to be dropped) or causes x^TCx to stay fixed and y_s to be dropped. If x_{r_1} is dropped, this theorem may be reapplied; on the other hand, if y_s drops, either initially or upon increase of y_r, Theorem 2 may be re-applied.

PROOF: Our proof is completely general; however, for convenience we will illustrate it on system (13) below. Let us suppose we had on some cycle a basis B and a basic feasible complementary solution with basic variables x_1, x_2, x_3, x_4, π_1, π_2, y_5 and the value of $y_5 = y_5^o < 0$. In this case, x_5 becomes a new basic variable and we assume that x_4 dropped out to form a new basis B'. In (13), the dot · indicates a column in the basis B and * indicates that the column P_5 associated with x_5 is a candidate to replace a vector of the basis B. Let the representation of P_5 in terms of the columns of the basis B be:

$$(12) \qquad P_1\alpha_1 + P_2\alpha_2 + P_3\alpha_3 + P_4\alpha_4 + P_6\alpha_6 + P_7\alpha_7 + \bar{P}_5\bar{\alpha}_5 = P_5$$

where $\overline{P}_5$ is the y_5 column in (13).

(13)

x_1	x_2	x_3	x_4	x_5	π_1	π_2	y_1	y_2	y_3	y_4	y_5	Const.
a_{11}	a_{12}	a_{13}	a_{14}	a_{15}								b_1
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}								b_2
c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	a_{11}	a_{21}	-1					0
c_{12}	c_{22}	c_{23}	c_{24}	c_{25}	a_{12}	a_{22}		-1				0
c_{13}	c_{23}	c_{33}	c_{34}	c_{35}	a_{13}	a_{23}			-1			0
c_{14}	c_{24}	c_{34}	c_{44}	c_{45}	a_{14}	a_{24}				-1		0
c_{15}	c_{25}	c_{35}	c_{45}	c_{55}	a_{15}	a_{25}					-1	0
.	.	.	.	*	.	.					.	basis B
.	.	.		.	.	.				*	.	basis B$'$

Let us now consider the representation of the y_4 column, $\overline{P}_4$, in terms of the basis B' where λ_i' are the weights on columns P_i associated with basic x_i, π_i and $\overline{\lambda}_i'$ are the weights on columns $\overline{P}_i$ associated with basic y_i ,

$$\text{(14a)} \qquad P_1\lambda_1' + P_2\lambda_2' + P_3\lambda_3' + P_5\lambda_5' + P_6\lambda_6' + P_7\lambda_7' + \overline{P}_5\overline{\lambda}_5' = \overline{P}_4 \quad .$$

We wish to show that $\lambda_5' \leqslant 0$. If $\lambda_5' < 0$, it is clear that an increase of y_4 will cause x_5 to increase and $x^T Cx$ to decrease as long as the value of $y_5 < 0$ in the basic solution. On the other hand, if $\lambda_5' = 0$, we shall show that y_5 will drop with no change in $x^T Cx$.

Let $[\lambda_i]$ be the representation of $\overline{P}_4$ in terms of the <u>prior basis</u> B, (i.e., before the introduction of x_5 in place of x_4),

$$\text{(14b)} \qquad P_1\lambda_1 + P_2\lambda_2 + P_3\lambda_3 + P_4\lambda_4 + P_6\lambda_6 + P_7\lambda_7 + \overline{P}_5\overline{\lambda}_5 = \overline{P}_4 \quad .$$

Then, setting $\lambda = (\lambda_1, \lambda_2, \lambda_3, \lambda_4)$, the first six rows of this representation yields (15) and (16)

$$(15) \qquad [a_{11} \quad a_{12} \quad a_{13} \quad a_{14}] \; \lambda^T = 0$$

$$[a_{21} \quad a_{22} \quad a_{23} \quad a_{24}] \; \lambda^T = 0$$

$$(16) \qquad \begin{bmatrix} c_{11} & c_{12} & c_{13} & c_{14} \\ c_{12} & c_{22} & c_{23} & c_{24} \\ c_{13} & c_{23} & c_{33} & c_{34} \\ c_{14} & c_{24} & c_{34} & c_{44} \end{bmatrix} \lambda^T + \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \end{bmatrix} \lambda_6 + \begin{bmatrix} a_{21} \\ a_{22} \\ a_{23} \\ a_{24} \end{bmatrix} \lambda_7 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ -1 \end{bmatrix}$$

Multiplying (16) by λ on the left and denoting the square matrix by C_4, yields, by (15), $\lambda C_4 \lambda^T = -\lambda_4$. Since $\lambda C_4 \lambda^T$ is positive semi-definite (C_4 is a principal minor of C), $\lambda C_4 \lambda^T \geqslant 0$ and $\lambda_4 \leqslant 0$ follows.

Case $\lambda_5' < 0$: Let us assume $\lambda_4 < 0$. We observe that in the representation (12) of P_5 in terms of B, the weight α_4 is positive (since x_4 decreased when x_5 increased). By eliminating P_4 from (12) and (14b) to obtain (14a), and noting $\alpha_4 > 0$, $\lambda_4 < 0$, it follows that $\lambda_5' = \lambda_4/\alpha_4 < 0$ (where λ_5' is the weight on P_5 in the representation of $\overline{P}_4$ in terms of B'). But $\lambda_5' < 0$ implies that the introduction of y_4 into the basic set for B' will increase x_5. Moreover, we may adopt the point of view, for the purpose of the proof, that it is the increase in x_5 that is "causing" the increase in y_4 (instead of the other way around), so that we are, in fact, repeating the situation just considered of increasing x_5 and adjusting the other "basic" variables, except here y_4 is in the basic set instead of x_4. It follows, therefore, as before, that an increase in x_5 decreases x^TCx as long as y_5 remains negative in value in the adjustment of the basic solution by the increase of x_5.

Case $\lambda'_5 = 0$: Let us now assume $\lambda_4 = 0$. We may set $\lambda_i = \lambda'_i$ because the representation of $\bar{P}_4$ is the same, whether in terms of B or B$'$; hence, $\lambda'_5 = 0$. In this case $\lambda C_4 \lambda^T = -\lambda'_5 = 0$; therefore, $C_4 \lambda^T = 0$ by a well known property of semi-definite forms. In this case $\lambda = 0$ must hold because $\lambda \neq 0$ implies a dependence of the first four columns of (15) and (16) which is impossible because then the square array of coefficients of (15) and (16), and in turn B, would be singular.

Setting $\lambda = 0$ in (16) and noting that at least one λ_i must be nonzero, $i=1,\ldots,7$, we see that there is a dependence between the rows of $[a_{ij}]$ for those columns x_j associated with the basic set, other than x_5. By forming a linear combination of the rows of A, we could therefore rewrite (for the purpose of the proof) the system so that top row has zero coefficients for these x_j. Thus

(17)

x_1	x_2	x_3	x_4	x_5	π'_1	π'_2	y_1	y_2	y_3	y_4	y_5	Const.
0	0	0	a'_{14}	a'_{15}								b'_1
a_{21}	a_{22}	a_{23}	a_{24}	a_{25}								b_2
c_{11}	c_{12}	c_{13}	c_{14}	c_{15}	0	a_{21}	-1					0
c_{12}	c_{22}	c_{23}	c_{24}	c_{25}	0	a_{22}		-1				0
c_{13}	c_{23}	c_{33}	c_{34}	c_{35}	0	a_{23}			-1			0
c_{14}	c_{24}	c_{34}	c_{44}	c_{45}	a'_{14}	a_{24}				-1		0
c_{15}	c_{25}	c_{35}	c_{45}	c_{55}	a'_{15}	a_{25}					-1	0
.	.	.		.	.	.				*	.	

Now, $a'_{14} \neq 0$ because B was nonsingular and $a'_{15} \neq 0$ because the same is true for B'. It is also obvious that the signs of a'_{14} and a'_{15} are the same for x_4 to decrease when x_5 increases. Note now that the π'_1 column is representable as a linear combination of the negative unit columns of y_4, y_5 (and, in a more general case than the example, the other negative unit columns of the basic y_j). Moreover it is clear that since a'_{14} and a'_{15} have the same sign, increasing y_4 from its zero value results in a positive change in y_5.

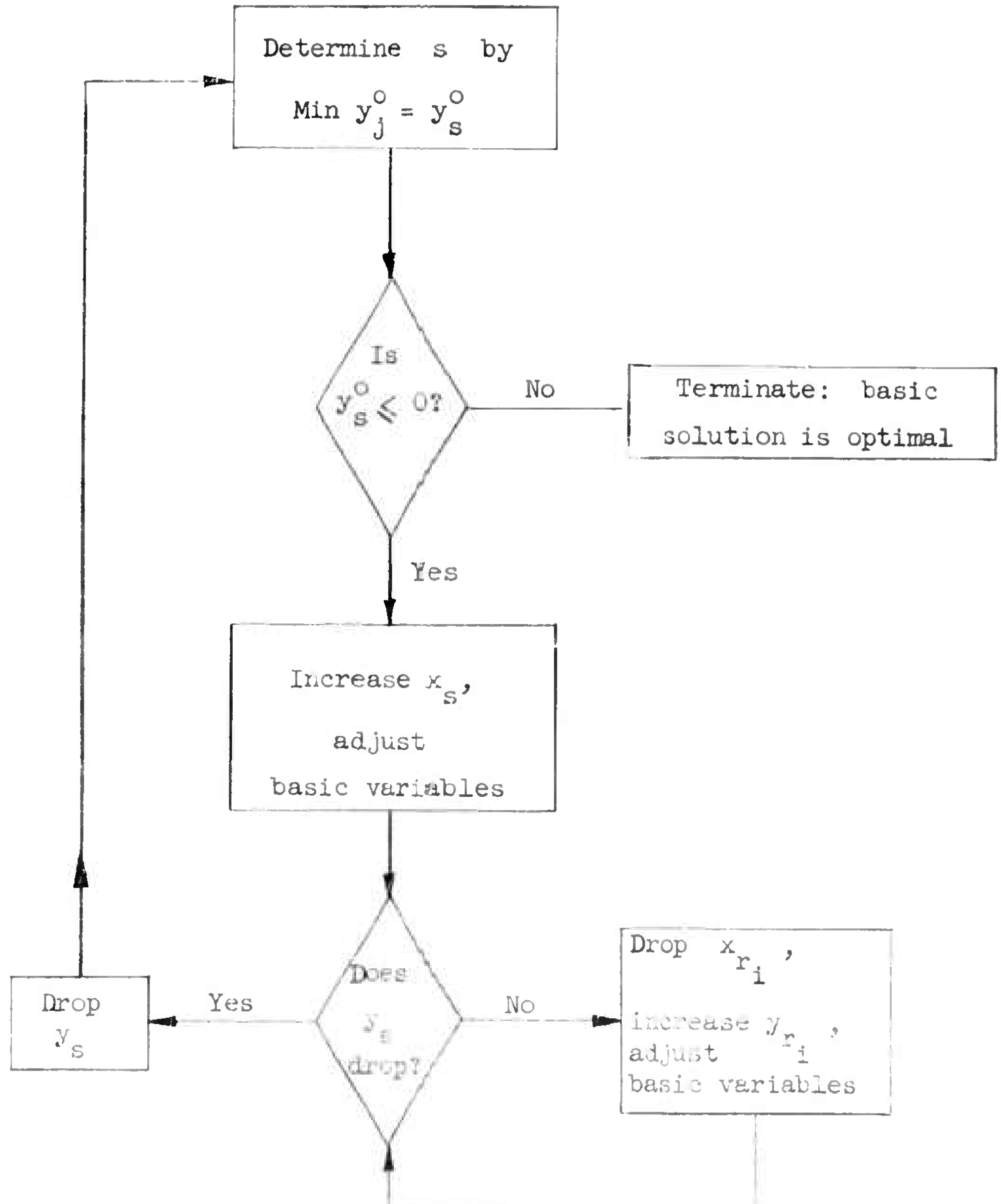
Since the y_j are not sign restricted, y_4 can be increased until y_5 is dropped out of the basic set at value zero because all x_j values are unaffected. Hence, in this shift of basis there is no change in the value of x^TCx; however, the introduction of y_4 into the basic set and dropping of y_5, gives rise to new basic set that satisfies the complementarity property. We may thus apply again Theorem 2 to reduce x^TCx.

THE QUADRATIC ALGORITHM:

STEP 1. Initiate: Let $Ax^0 = b$ be a basic feasible solution for $Ax = b$, $x \geqslant 0$, with basic variables $x_{j_1}, x_{j_2}, \ldots, x_{j_m}$; choose for the initial set of basic variables for the enlarged problem: (a) these x_{j_i}, (b) the complements, y_j, of the nonbasic x_j; and (c) the set π_i. (Their coefficient matrix is nonsingular.)

STEP 2. For the values of y_j^0 of the basic solution, determine $\text{Min}\ y_j^0 = y_s^0$. If $y_s^0 > 0$ terminate; the solution is optimal. If $y_s^0 < 0$ introduce into basic set x_s; if y_s drops from basic set, repeat Step 2. Otherwise if x_r drops,

STEP 3. Introduce y_r into basic set. If y_s drops, return to Step 2; otherwise, if some x_{r_1} drops, repeat Step 3 with r_1 playing the role of r.



THEOREM 4. The iterative process is finite.

PROOF: The number of possible basic sets is finite. Each one generated by the process is different because of the decreases in xCx. But this means the cyclic process must terminate.

CONCLUSION:

Formula (10) is the analog for quadratic programs of the familiar adjusted objective function obtained by elimination of the basic variables in linear programming. [For general convex objectives, it appears to be a natural take-off for a quadratic fit.] If the coefficients y_j^o of the x_j are non-negative, the solution is optimal. If not, a new basic solution for system (11) is obtained by increasing x_s corresponding to $y_s^o = \text{Min } y_j^o$. Either y_s drops out as basic variable or y_s drops after a sequence of replacements of basic variables x_r by their correspondents y_r. With the latter provision for a decrease in the dimensionality of the solution, the algorithm may be viewed as a direct extension of the regular simplex method to quadratic programs (in contrast, the algorithms of Wolfe and Markowitz may be viewed as parametric extensions).

BIBLIOGRAPHY

1. Barankin, E. W. and R. Dorfman, On Quadratic Programming, University of California Press, Berkeley: 1958.

2. Beale, E.M.L., "On Quadratic Programming," Naval Research Logistics Quarterly 6, Sept. 1959, pp 227-243.

3. Dorfman, R., Application of Linear Programming to the Theory of the Firm, University of California Press, Berkeley: 1951.

4. Frank, M. and P. Wolfe, "An Algorithm for Quadratic Programming," Naval Research Logistics Quarterly, March and June 1956, pp 95-110.

5. Houthakker, H. S., "The Capacity Method of Quadratic Programming," Proceedings, RAND Symposium of Math. Programming, The RAND Corp., Santa Monica: March 1959.

6. Kuhn, H. W. and A. W. Tucker, "Non-Linear Programming," Proceedings of the Second Berkeley Symposium on Math. Stat. and Probability, University of California Press, Berkeley: 1951, pp 481-92.

7. Markowitz, H., "The Optimization of a Quadratic Function Subject to Linear Constraints," Naval Logistics Quarterly, March and June 1956, pp 111-133; also, Appendix, Portfolio Selection: Efficient Diversification of Investments, John Wiley and Sons, Inc., New York: 1959.

8. White, W. B., S. M. Johnson and G. B. Dantzig, "Chemical Equilibrium in Complex Mixtures," The Journal of Chemical Physics, vol. 28, no. 5, May 1958, pp 751-755; also, "A Linear Programming Approach to the Chemical Equilibrium Problem," Management Science, vol. 5, no. 1, October 1958.

9. Wolfe, P., "The Simplex Method for Quadratic Programming," Econometrica, vol. 27, no. 3, July 1959.

10. Zoutendijk, G., Methods of Feasible Directions, D. Van Nostrand Co., Inc., New York: 1960.